SF1901: Övningshäfte

13 oktober 2013

Uppgifterna under rubriken *Övning* kommer att gås igenom under övningstillfällena. Uppgifterna under rubriken *Hemtal* är starkt rekommenderade och motsvarar nivån på tentamen. Om de uppgifterna upplevs som för svåra finns det enklare uppgifter i kursboken. De ligger sist på varje övning under rubriken *Blom*.

12 Övning 12: Hypotesprövning

Uppgift 12.1 *Per spelar på en spelautomat som ger vinst med den okända sannolikheten p. Antalet spel X till och med första vinsten har då sannolikhetsfunktionen*

$$p_X(k) = p(1-p)^{k-1} \quad \text{för } k = 1, 2, \ldots.$$

Det påstås att $p = 0.2$ men Per betvivlar att p är så stort och vill pröva hypotesen $H_0 : p = 0.2$ mot $H_1 : p < 0.2$. Kan han på signifikansnivån högst 0.10 förkasta H_0 om han finner att han förlorar de tio första spelen och vinner det elfte? Använd P-värdesmetoden.

Uppgift 12.2 *En tillverkare av stålrör vill undersöka om en viss ytbehandling har någon korrosionsminskande effekt. Man har grävt ned 18 obehandlade och 18 behandlade rör parvis med ett behandlat och ett obehandlat rör i varje par. Man antar sedan att de båda rören i ett par befinner sig i samma miljö.*

Efter en tid skall man gräva upp rören och se efter i hur många par som det behandlade röret korroderat mest, säg i x par (vi förutsätter att man alltid kan se någon skillnad).

Som nollhypotes har valts att rören korroderar lika mycket, dvs att sannolikheten p att ett behandlat rör korroderat mer än ett obehandlat är 0.5. Som signifikanstest har valts att förkasta hypotesen om $x \leq 5$.

a) Bestäm testets signifikansnivå.

b) Vad blir slutsatsen om $x = 3$ respektive om $x = 5$ respektive om $x = 15$?

c) Beräkna styrkan mot alternativet $p = 0.25$.

Uppgift 12.3 *En läkemedelstillverkare använder ibland en viss livsmedelsfärg. Man vill veta hur färgen påverkar utseendet hos det framställda läkemedlet. Ur tillverkningen tar man därför på måfå tio förpackningar och mäter grumligheten i innehållet efter en tids lagring. Resultat:*

3.9 4.1 4.4 4.0 3.8 4.0 3.9 4.3 4.2 4.4

Utan färgtillsats brukar grumligheten vara i medeltal 4.0. Man undrar nu om resultaten tyder på att grumligheten ökar. Modell: Materialet anses vara ett slumpmässigt stickprov från $N(\mu, 0.2)$. Pröva hypotesen $H_0 : \mu = 4.0$ mot $H_1 : \mu > 4.0$ med ett test på nivån 0.05.

Uppgift 12.4 *Fortsättning från Uppgift 12.3. Om μ är det rätta värdet, vilken fördelning har den s.v. som teststorheten är en observation av? Beräkna styrkefunktionen för testet, dvs bestäm $\mathrm{P}(H_0$ förkastas), om μ är det rätta värdet. Vilken styrka har testet för $\mu = 3.8$? För $\mu = 4.3$?*

Uppgift 12.5 *En forskare har bildat en ny legering och teoretiskt beräknat dess smältpunkt till 1050°. För att kontrollera resultatet har han mätt smältpunkten hos 10 prover av legeringen och erhållit följande mätvärden:*

1054.8	1052.9	1051.0	1049.8	1051.6
1047.9	1051.8	1048.5	1050.2	1050.7

Variationerna i mätvärden beror på imperfektioner hos termometern. Erfarenhet från tidigare försök ger att man kan anta att mätfelen är oberoende och normalfördelade med väntevärde noll och standardavvikelse 2.3.

a) *Testa hypotesen att smältpunkten μ är 1050° på nivån 5%. Som mothypotes tas att smältpunkten är skild från 1050°.*

b) *Bestäm testets styrkefunktion och beräkna styrkan för alternativen $\mu = 1051$ och $mu = 1053$.*

12.1 Hemtal 12

Uppgift 12.6 *På misstänkta rattfyllerister gjordes tidigare tre bestämningar av alkoholhalten i blodet. Resultatet x_1, x_2 och x_3 antogs utgöra ett slumpmässigt stickprov från $N(\mu, 0.05)$ där μ är det verkliga alkoholhalten (promille). Om $\mu > 0.5$ har personen gjort sig skyldig till rattonykterhet. Låt oss anta att den domstol som skulle döma, tog hänsyn till osäkerheten i blodproven genom att beräkna aritmetiska medelvärdet $\bar{x}$ av de tre analysresultaten och därefter förklara personen skyldig till rattonykterhet om*

$$\bar{x} > 0.5 + \lambda_{0.01} \cdot 0.05/\sqrt{3}$$

och oskyldig annars. Med statistisk terminologi kan man säga att domstolen testar $H_0 : \mu = 0.5$ mot $H_1 : \mu > 0.5$ på signifikansnivån 0.01. Vilka av följande påståenden ger en någorlunda korrekt beskrivning av vad som hände i det långa loppet?

1. *Högst 1% av alla frikända var skyldiga.*
2. *Högst 1% av alla oskyldiga blev dömda.*
3. *Högst 1% av alla skyldiga blev frikända.*
4. *Högst 1% av alla dömda var oskyldiga.*

Uppgift 12.7 *Fortsättning från Uppgift 12.5. Antag att standardavvikelsen inte är känd i förväg. Testa hypotesen att smältpunkten är 1050° på nivån 5%. Som mothypotes tas att smältpunkten är skild från 1050°.*

Uppgift 12.8 *Vid kvicksilverundersökning av gäddor i en insjö har man bestämt kvicksilverhalten i 10 fångade gäddor av viss storlek. Resultat (mg/kg):*

0.8   1.6   0.9   0.8   1.2   0.4   0.7   1.0   1.2   1.1

Modell: Halten i gäddor av den aktuella storleken varierar som en s.v. $X \in N(\mu,\sigma)$.

a) *Kan man med de erhållna resultaten på signifikansnivån 0.05 förkasta* $H_0 : \mu = 0.9$ *mot* $H_1 : \mu > 0.9$?

b) *Kan man på signifikansnivån 0.05 förkasta* $H_0 : \mu = 1.1$ *mot* $H_1 : \mu < 1.1$?

Uppgift 12.9 *Man har anledning att förmoda att molekylvikterna för två kemiska föreningar A och B är lika. För att undersöka detta har man gjort 6 molekylviktsbestämningar på A och 8 på B. Den använda mätmetoden ger ett mätfel som är* $N(0,\sigma)$. *Mätfelen i olika mätningar kan anses vara oberoende. Följande värden erhölls:*

Molekylvikt för A	174.18	174.30	174.23	174.29	174.36	174.25		
Molekylvikt för B	174.19	174.40	174.20	174.35	174.32	174.14	174.27	174.34

a) *Konstruera ett test på nivån 5% av hypotesen* H_0 : *A och B har samma molekylvikt, mot alternativet* H_1: *A och B har olika molekylvikt.*

b) *Gör numeriska beräkningar och ange testets utfall.*

Uppgift 12.10 *Åtta personer mäter sin längd (cm) morgon och kväll. Resultat:*

Person	1	2	3	4	5	6	7	8
Morgon	172	168	180	181	160	163	165	177
Kväll	172	167	177	179	159	161	166	175

Skillnaderna mellan morgon- och kvällsvärdena antas vara ett slumpmässigt stickprov från $N(\mu,\sigma)$. *Pröva hypotesen* $H_0 : \mu = 0$ *mot* $H_1 : \mu \neq 0$ *på signifikansnivån 0.05.*

Uppgift 12.11 *Vid tillverkning av magnecyl varierar tablettvikten som en s.v. med väntevärdet* μ *och standardavvikelsen* $\sigma = 0.02$. *För kontroll väger man 35 tabletter och beräknar som punktskattning av* μ *de 35 tabletternas medelvikt* $\bar{x}$, *vilken blev 0.69 g. Pröva hypotesen* $H_0 : \mu = 0.65$ *mot* $H_1 : \mu \neq 0.65$ *med ett test med den approximativa signifikansnivån 0.05. (Observera: ingen förutsättning om normalfördelning hos tablettvikterna.)*

Blom: 13.1, 13.2, 13.3

13 Övning 13: χ^2-test

Uppgift 13.1 *Vid ett genetiskt försök studerade man den avkomma som erhölls vid korsning mellan två olika typer av marsvin. Av 87 ungar var 43 röda, 10 svarta och 34 vita. Enligt den genetiska modellen skulle dessa färger förekomma i proportionen 9:3:4, dvs med sannolikheterna 9/16, 3/16 och 4/16. Skall hypotesen om en sådan fördelning förkastas på nivån 5%?*

Uppgift 13.2 *Vid en trafikräkning på en av Sveriges länsvägar räknades under 81 vardagar antalet bilar som passerade en viss plats i ena riktningen kl. 9.00-9.10. Följande material erhölls:*

Antal bilar	0	1	2	3	4	5	6
Antal dagar	14	12	25	16	10	3	1

Testa om antalet bilar per tiominutersperiod kan anses vara Poissonfördelat. Välj nivån 5%.

Uppgift 13.3 *Ur var och en av tre stora populationer P_1, P_2 och P_3 av personer valde man slumpmässigt ut ett stickprov och klassificerade de uttagna personerna enligt följande tabell:*

	män	kvinnor
P_1	46	54
P_2	78	72
P_3	143	107

Pröva på signifikansnivån 5% hypotesen att könsfördelningen är densamma i de tre populationerna.

Uppgift 13.4 *Från 500 olyckor på landsväg har man samlat in data enligt nedanstående tabell:*

Skador	Antal olyckor då säkerhetsbälte	
	användes	icke användes
Inga eller lätta personskador	102	143
Svåra personskador	58	198

Pröva på signifikansnivån 5% om användningen av säkerhetsbälte påverkar skadetypen.

13.1 Hemtal 13

Uppgift 13.5 *Den s.v. X antar värdena 0, 1, 2, 3. Man gjorde 4096 oberoende observationer av X och fick*

Observation	0	1	2	3
Antal	1764	1692	552	88

Pröva på signifikansnivån 1% hypotesen att $X \in Bin(3,1/4)$.

Uppgift 13.6 *Försäkringsbolaget, FörSäkra AB, vill undersöka om risken att råka ut för en trafikolycka är oberoende av förarens ålder. Man valde därför slumpmässigt ut 1000 av sina bilförsäkringar och registrerade personens ålder och antalet trafikolyckor under tiden mellan 2012-01-01 och 2012-12-31. Resultatet:*

Antal olyckor	18-21 år	22-30 år	31-40 år	41-50 år	51-70 år
0	162	251	179	118	30
1	49	41	22	24	4
2 eller fler	39	28	9	28	16

Undersök med ett lämpligt statistiskt test, på nivån 1% om risken att råka ut för olycka är beroende av åldern? Ett tydligt svar bör framgå. Svaret och den tillämpade statistiska metoden bör motiveras.

Uppgift 13.7 *Vid en undersökning rörande trafikskadade i Västergötland under tiden mellan 2008-09-01 och 2010-02-28 har man bland annat studerat hur utsatta olika trafikantkategorier är för skallskador. Följande material insamlades för totalt 695 skallskadade personer:*

	Bilister	Cyklister, mopedister, motorcyklister	Fotgängare
Ej medvetslös	122	80	28
Medvetslös högst 30 min	179	163	62
Medvetslös mer än 30 min	25	22	14

Undersök med ett lämpligt statistiskt test, på nivån 5% om de tre trafikantkategorierna uppvisar någon skillnad vad gäller skallskadornas svårighetsgrad. Ange tydligt vilka de uppställda hypoteserna och slutsatsen är.

Uppgift 13.8 *Antalet tryckfel i första upplagan av en bok ges av följande data:*

Antal tryckfel på en sida	0	1	2	>2
Antal sidor	249	42	9	0

Testa om antalet tryckfel på en sida kan anses vara Poissonfördelat. Välj nivån 5%.

Blom: 13.29, 13.32, 13.13, 13.20, 13.25

14 Övning 14: Maximum-likelihood-metoden, Minsta-kvadrat-metoden

Uppgift 14.1 *Den diskreta s.v.* X *har sannolikhetsfunktionen* $p_X(k) = \theta(1-\theta)^{k-1}$ *för* $k = 1, 2, \dots,$ *där* $0 < \theta < 1$*. Man har ett slumpmässigt stickprov 4, 5, 4, 6, 4, 1 från denna fördelning. Bestäm ML-skattningen av* θ*.*

Uppgift 14.2 *Antalet fartyg som under ett tidsintervall av längden* t *(enhet: minut) passerar Helsingborg på väg söderut genom Öresund anses vara* $Po(\lambda t)$*. Antalet fartyg i skilda intervall anses oberoende. En person vill uppskatta* λ *och räknar under tre olika tidsintervall antalet passerande fartyg.*

Observationstid	30	30	40
Antal fartyg	10	12	18

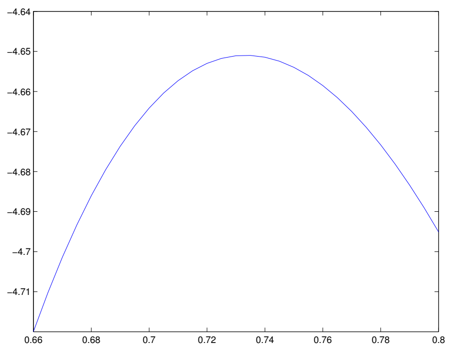
a) Ange ML-skattningen av λ.

b) Hur stor är standardavvikelsen för ML-skattningen?

Uppgift 14.3 *Väntetiden [s] från start till det fjärde anropet till en server beskrivs som en stokastisk variabel* X *som har täthetsfunktionen*

$$f_X(x) = \begin{cases} \frac{x^3}{6\theta^4} e^{-x/\theta}, & x > 0, \\ 0, & \text{för övrigt.} \end{cases}$$

Vi har tre oberoende observationer av dessa väntetider, $x_1 = 2.3$*,* $x_2 = 1.9$*,* $x_3 = 4.6$*.*



Figur 1: Loglikelihoodfunktionen $l(\theta)$ i Uppgift 14.3.

a) *I Figur 1 har loglikelihoodfunktionen* $l(\theta)$*, för de givna observationerna, plottats med* θ *i* $[0.66, 0.9]$*. Ge med ögonmått ett intervall (högst av längd 0.05), där maximum-likelihood-skattningen av* θ *borde ligga.*

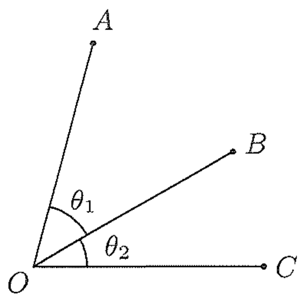
b) *Bestäm en punktskattning av* θ *på basis av dessa observationer med maximum-likelihood-metoden. Kontrollera med avseende på svaret i a).*

Uppgift 14.4 *Vissa problem i partikelfysik ger upphov till följande sannolikhetstäthet*

$$f(x) = \begin{cases} \frac{1}{2}(1+\theta x), & -1 \leq x \leq 1, \\ 0, & \textit{för övrigt,} \end{cases}$$

där $-1 \leq \theta \leq 1$ *är en okänd parameter. Antag att* $X_1, \ldots, X_n$ *är oberoende stokastiska variabler som alla har denna fördelning. Bestäm MK-skattningen* $\theta^{\star}_{MK}$ *av* θ.

Uppgift 14.5 *Man har gjort tre oberoende mätningar av vinkeln* AOC *och erhållit mätvärdena* x_1, x_2, x_3 *samt gjort två oberoende mätningar av vinkeln* AOB*, mätvärdena* x_4, x_5 *(se Figur 2). Vid varje vinkelmätning fås ett slumpmässigt fel som har väntevärdet 0 och standardavvikelsen* σ. *Antag att vinklarna är* $\theta_1 + \theta_2$ *respektive* θ_1 *grader.*



Figur 2: Vinklarna AOB och AOC i Uppgift 14.5.

a) Bestäm MK-skattningen av θ_1 *och* θ_2.

b) Undersök om skattningarna är väntevärdesriktiga.

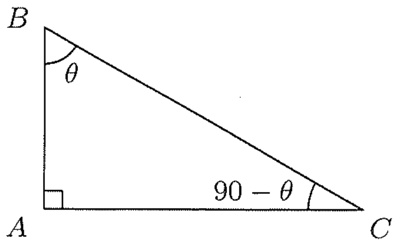
c) Beräkna variansen för skattningarna.

14.1 Hemtal 14

Uppgift 14.6 *Man har ett slumpmässigt stickprov* $x_1, \ldots, x_n$ *från en fördelning med täthetsfunktionen* $f_X(x) = \theta x^{\theta-1}$ *för* $0 < x < 1$. *Ange ML-skattningen av* θ.

Uppgift 14.7 *Vinklarna vid B och C i den rätvinkliga triangeln mättes (se Figur 3). Resultat:*

B:	61.2	61.4	61.1	
C:	28.9	28.9	28.9	28.8



Figur 3: Vinklarna B och C i Uppgift 14.7.

Mätningarna vid B och C kan anses vara observationer av stokastiska variabler med väntevärde θ respektive $90 - \theta$ och med samma varians σ^2.

a) Skatta med hjälp av MK-metoden vinklarna vid B och C.

b) Undersök om skattningarna är väntevärdesriktiga.

c) Beräkna standardavvikelsen för skattningarna då $\sigma = 0.1$.

Uppgift 14.8 *Två forskare A och B har genom en stickprovsundersökning skattat andelen färgblinda, p, i en stor population av män. A har valt ut 1000 personer och funnit 77 färgblinda, medan B har undersökt 2000 och funnit 124 färgblinda. Bestäm ML-skattningen av p.*

Uppgift 14.9 *För att bestämma arean θ av en kvadrat har man gjort oberoende mätningar utan systematiska fel av kvadratens sida och därvid fått värdena $x_1, \ldots, x_n$. Dessa antages vara observationer från $N(\mu, \sigma)$, där μ är okänd och σ är känd. (Uppenbarligen är $\theta = \mu^2$.)*

a) Ange ML-skattningen av θ.

b) Ange den korrigerade ML-skattningen av θ.

Uppgift 14.10 *(Klurig) Ett mätfel av glappkaraktär har likformig fördelning $[-\theta,\theta]$. Antag att man vid mätning av kända storheter fått mätfelen $x_1, \ldots, x_n$. Härled ML-skattningen av θ.*

15 Övning 15: Reserv

Svar

Övning 12

12.1 Efter $P = 0.107 > 0.10$ kan H_0 ej förkastas.
12.2 a) $\alpha = 0.04813$, b) $x = 3$; förkasta H_0, $x = 7$; förkasta ej H_0, $x = 15$; förkasta ej H_0, c) $h(0.25) = 0.71745$.
12.3 Då $z = 1.58 < \lambda_{0.05}$ kan H_0 ej förkastas på signifikansnivån 0.05.
12.4 $h(3.8) \approx 0$, $h(4.3) \approx 0.9990$.
12.5 a) Då $z = 1.2649 < \lambda_{0.025}$ kan H_0 ej förkastas på signifikansnivån 0.05, b) $h(1051) = 0.28$, $h(1053) = 0.98$.
12.6 Endast påstående 2.
12.7 Då $t = 1.4345 < t_{0.025}(9)$ så H_0 kan ej förkastas på signifikansnivån 0.05.
12.8 a) $t = 0.67 < t_{0.05}(9)$; H_0 kan ej förkastas, b) $t = -1.24 > -t_{0.05}(9)$; H_0 kan ej förkastas.
12.9 a) $H_0 : \mu_A = \mu_B$ vs. $H_1 : \mu_A \neq \mu_B$, nivå $\alpha = 0.05$. Förkasta H_0 om $|\frac{(\bar{x}-\bar{y})-(\mu_A-\mu_B)}{s\sqrt{1/n_1+1/n_2}}| > t_{0.025}(6+8-2) = 2.18$, där $s^2 = \frac{(n_1-1)s_x^2+(n_2-1)s_y^2}{(n_1-1)+(n_2-1)}$. b) $s = 0.0808$ ger $0.1822 < t_{0.025}(12)$; förkasta ej H_0.
12.10 $t = 2.76 > t_{0.025}(7)$; H_0 förkastas.
12.11 $z = 11.8$; man får $|11.8| > \lambda_{0.025}$; H_0 förkastas.

Övning 13

13.1 $Q_{obs} = 10.0 > \chi^2_{0.05}(2) = 5.99$. Förkasta hypotesen.
13.2 H_0: antalet bilar per tiominutersperiod är Poissonfördelat. $Q_{obs} = 6.0 < \chi^2_{0.05}(3) = 7.81$. Förkasta ej hypotesen.
13.3 $Q_{obs} = 3.77 < \chi^2_{0.05}(2) = 5.99$; H_0 kan ej förkastas.
13.4 H_0: användning av säkerhetsbälte påverkar ej skadetypen. $Q_{obs} = 20.00 > \chi^2_{0.05}(1) = 3.84$. Förkasta hypotesen.
13.5 $Q_{obs} = 11.5 > \chi^2_{0.01}(3) = 11.3$. H_0 förkastas.
13.6 H_0: Risken att råka ut för olycka och ålder är oberoende. $Q_{obs} = 53.6043 > \chi^2_{0.01}(8)$; H_0 förkastas på nivån 1%.
13.7 H_0: Skallskadornas svårighetsgrad är likvärdig. $Q_{obs} = 7.9813 < \chi^2_{0.05}(4)$; H_0 kan ej förkastas på nivån 5%.
13.8 H_0: antalet tryckfel per sida är Po(μ). $Q_{obs} = 3.75 < \chi^2_{0.05}(1) = 3.84$. Förkasta ej hypotesen.

Övning 14

14.1 $\theta^*_{obs} = 1/\bar{x} = 1/4$.
14.2 a) $\lambda^*_{obs} = \frac{\sum_{i=1}^n x_i}{\sum_{i=1}^n t_i} = 0.4$, b) $D(\lambda^*) = \lambda/10$.
14.3 a) ML-skattningen ligger inom $[0.72, 0.74]$, b) $\theta^*_{obs} = \frac{\sum_{i=1}^3 x_k}{12} = 0.7333$.
14.4 $\theta^*_{obs} = \frac{3}{n}\sum_{i=1}^n x_i$.
14.5 a) $(\theta_1)^*_{obs} = \frac{1}{2}(x_4 + x_5)$, $(\theta_2)^*_{obs} = \frac{1}{6}(2x_1 + 2x_2 + 2x_3 - 3x_4 - 3x_5)$, b) Skattningarna är väntevärdesriktiga, c) $\mathrm{Var}(\theta_1^*) = \sigma^2/2$, $\mathrm{Var}(\theta_2^*) = 5\sigma^2/6$.
14.6 ML-skattningen är $\theta^*_{obs} = -n/\sum \ln x_i$.
14.7 a) 61.17 respektive 28.82, b) Skattningarna är väntevärdesriktiga, c) För bägge skattningarna är standardavvikelsen $\sigma/\sqrt{7} = 0.038$.

14.8 ML-skattningen är $p^*_{obs} = 0.067$.
14.9 a) $\bar{x}^2$, b) $\bar{x}^2 - \sigma^2/n$.
14.10 $\theta^*_{obs} = \max(|x_1|, \ldots, |x_n|)$.